Measurement of Adrenal Steroid Hormones

Buttler, R.M.

2016

document version
Publisher's PDF, also known as Version of record

Chapter VI

Comparison of eight routine unpublished LC-MS/MS methods for the simultaneous measurement of testosterone and androstenedione in serum

Rahel M. Büttler[1], Frans Martens[1], Mariëtte T. Ackermans[2], Andrew Davison[3], Antonius E. van Herwaarden[4], Linda Kortz[5], Johannes G. Krabbe[6], Eef G.W. Lentjes[7], Charlotte Syme[8], Rachel Webster[9], Marinus A. Blankenstein[1], Annemieke C. Heijboer[1]

[1]Department of Clinical Chemistry, Endocrine Laboratory, VU University Medical Center, Amsterdam, The Netherlands. [2]Department of Clinical Chemistry, Laboratory of Endocrinology, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands. [3] Department of Clinical Biochemistry and Metabolic Medicine, Liverpool Clinical Laboratories, Royal Liverpool and Broadgreen University Hospitals Trust, Liverpool, United Kingdom. [4] Department of Laboratory Medicine, Radboud University Medical Centre, Nijmegen, The Netherlands. [5] MVZ Wagnerstibbe für Laboratoriumsmedizin, Gynäkologie, Humangenetik und Pathologie GmbH, Göttingen, Germany. [6] Department of Clinical Chemistry and Laboratory Medicine, Medical Spectrum Twente, Medlon BV, Enschede, The Netherlands. [7]Department of Clinical Chemistry & Haematology, University Medical Center, Utrecht, the Netherlands. [8] Department of Clinical Biochemistry, Glasgow Royal Infirmary, Glasgow, United Kingdom. [9] Department of Clinical Biochemistry - University Hospitals Birmingham NHS Foundation Trust Queen Elizabeth Hospital, Birmingham, United Kingdom.

Clin Chim Acta 2016; 454:112-8

Abstract

Background Liquid-chromatography tandem mass spectrometry (LC-MS/MS) has become the method of choice in steroid hormone measurement. However, little information on the mutual agreement of LC-MS/MS methods is available. We compared eight routine LC-MS/MS methods for the simultaneous measurement of testosterone and androstenedione.

Methods Sixty random serum samples from male and female volunteers were analyzed in duplicate by eight routine LC-MS/MS methods. We performed Passing-Bablok regression analyses and calculated Pearson's correlation coefficients to assess the agreement of the methods investigated with one published method known to be accurate. An intra-assay CV of each method was calculated from duplicate results, recoveries for each method were calculated from six spiked samples. Furthermore, a CV between the investigated methods was calculated.

Results The concentrations ranged from 0.05-1.26 nmol/L, 6.15-24.44 nmol/L and 0.15-4.78 nmol/L for testosterone in females, testosterone in males and androstenedione, respectively. The intra-assay CVs were between 3.7-16.0%, 0.9-5.2% and 1.2-9.5% for testosterone in females, testosterone in males and androstenedione, respectively. The slopes of the regression lines ranged between 0.90-1.25, 0.87-1.24 and 0.94-1.31 for testosterone concentrations in females, all testosterone values and androstenedione, respectively. The inter-method CVs were 24%, 14% and 29% for testosterone for concentrations in females and males and androstenedione, respectively. These compare unfavorably to the variation found earlier in published methods.

Conclusion Although most routine LC-MS/MS methods investigated here showed a reasonable agreement, some of the assays showed a high variation. The observed differences in standardization should be taken into account when applying reference values, or should, preferably, be solved.

Introduction

Accurate measurement of hormone concentrations is vital for clinical endocrinology as well as endocrine research. In recent years, more and more attention is paid on the reliability of hormone analysis, especially steroid hormone measurement. Taieb et al. showed that commonly used immunoassays for serum testosterone are not able to reliably measure the low testosterone concentrations in females and children [1]. The editorial, accompanying Taieb et al.'s publication, stated that these immunoassays are comparable to or even worse than a random number generator in estimating serum testosterone concentrations in females [2]. Since this publication in 2003 [1], an extensive debate on the quality of testosterone assays emerged and resulted in the Endocrine Society's Position Statement and the subsequent Consensus statement on testosterone assays, stating that the accuracy of testosterone measurements needs improvement [3;4].

Much effort has since been put in the development of more reliable assays for sex steroids. This resulted in second generation testosterone immunoassays, some of which were found to be more accurate than their predecessors [1;5;6]. Next to this, liquid-chromatography tandem mass spectrometry (LC-MS/MS) is more and more commonly used in steroid hormone analysis to avoid cross reactivity, which is one of the issues in immunoassays [7]. The urgent need for reliable assays as well as the increasing use of LC-MS/MS has prompted the editorial board of the Journal of Clinical Endocrinology and Metabolism (JCEM) to state that from 2015 on *'manuscripts reporting sex steroid assays as important endpoints must use MS-based assays'* and *'it is anticipated that this* [requirement] *will be extended to adrenal steroids and vitamin D in the near future'* [8]. Although we fully agree that reliable steroid hormone assays are required and we understand the superiority of LC-MS/MS above immunoassays, we believe that this statement is ambitious, especially because information on the mutual agreement of LC-MS/MS assays is still limited.

Thienpont et al. showed that LC-MS/MS assays for testosterone may agree well with each other and with a reference method and have a low imprecision [9]. However, others showed a less strong agreement and a much higher variation in the investigated LC-MS/MS methods for testosterone [10;11]. Data on the agreement of LC-MS/MS assays for androstenedione are even scarcer [12]. We recently compared seven published LC-MS/MS methods for the simultaneous measurement of testosterone, androstenedione and dehydroepiandrosterone (DHEA) and found that these assays agreed reasonably well. In addition, these published LC-MS/MS assays showed a clearly lower inter-method variation than currently used immunoassays for serum testosterone [13]. These findings support the JCEM statement. However, based upon earlier findings [10;11], it is questionable whether the reported findings apply to unpublished LC-MS/MS methods developed for routine diagnostic use as well. For this reason, we investigated the imprecision, trueness and agreement among eight routine and unpublished LC-MS/MS methods for the simultaneous measurement of testosterone and androstenedione.

Materials and Methods

Samples Sixty random serum samples were obtained from adult volunteers (men and women) presenting at the outpatient clinic of the VU University Medical Center Amsterdam for diagnostic venipuncture in March 2015. There were no further selection criteria and all subjects provided written informed consent. All samples were anonymized and handled identically. After centrifugation the samples were aliquotted and frozen at -20 °C. The samples were sent frozen to the participating laboratories and were kept frozen until analysis. We spiked three female samples with 1 nmol/L testosterone (Riedel de Haën (lot: 5117X), Buchs, Switzerland) and 10 nmol/L androstenedione (Steraloids (lot A6030-100, batch L1712) Newport, RI) and three male samples with 10 nmol/L testosterone and 4 nmol/L androstenedione.

Method comparison Eight routine, unpublished methods for the simultaneous measurement of testosterone and androstenedione were included for this comparison study. One published LC-MS/MS was included [12;13]. The methods were randomly coded Method A to Method J, with method A being the published method. Technical details of the LC-MS/MS methods in this study are shown in Table 1. In summary, the methods used between 25 and 500 µL serum for singular analysis. Sample preparation consisted of internal standard addition and one or more of the following sample preparation methods: liquid-liquid extraction, protein precipitation using acetonitrile, solid phase extraction and supported liquid extraction. The calibration ranges, Lower Limits of Quantitation (LLOQ) and internal standards used of all methods are shown in Table 2. Some of the methods measure several other steroids besides testosterone and androstenedione. However, in this study only the testosterone and androstenedione results were compared. All samples were analyzed in duplicate for testosterone and androstenedione by each of the investigated methods according to the standard procedures concerning calibration and quality control in each of the laboratories. Duplicate measurements were performed in one batch to allow calculation of the intra-assay coefficient of variation (CV) per method. We compared each method to method A, a published

method, shown to agree well with six published LC-MS/MS methods for testosterone and androstenedione and to be indirectly comparable with the testosterone reference method [12;13]. All samples were measured in duplicate on two different days using method A. The mean concentration per sample (the mean of the two duplicates) measured by method A was used for further analysis.

	Sample volume (µL)	Instrument	Analytical column	Sample preparation
Method A	100	Waters Acquity 2D-UPLC and Xevo TQ-s MS/MS	Waters Acquity UPLC BEH C4 Column, 1.7 µm, 2.1x50 mm and Phenomenex Kinetex Fluorophenyl 1.7 µm, 2.1x100 mm	Liquid-liquid extraction using hexane-ether (4:1)
Method B	100	Agilent Technologies Agilent 1290 Bin UPLC and 6490 Triple Quadrupole	Waters Acquity UPLC BEH C18 Column, 1.7 µm, 2.1x50 mm	Protein precipitation using acetonitrile + 0.1% formic acid and SPE using Waters Oasis HLB 1cc Extraction Cartridges
Method C	25	Waters Acquity UPLC and Xevo TQS	Waters Acquity HSS T3 column, 1.8 µm, 2.1x100 mm,	Protein precipitation using acetonitrile and Porvair Sciences P3 filtration plate
Method D	100	Waters Acquity LC and Water Quattro Premier XE MS/MS	Phenomenex Kinetex XB-C18, 5µm, 2.1x50 mm	Liquid-liquid extraction using diethyl ether
Method E	200	Waters Alliance 2795 LC and Quattro Premier XE- MS/MS	Waters C18 Atlantis column 3.0µ, 2.1 x 50mm	Liquid liquid extraction using MTBE
Method F	100	Waters Acquity LC and Xevo TQ-s MS/MS	Acquity UPLC BEH C18, 1.7 µm, 2.1 x 50 mm	Extraction with ISOLUTE® Supported Liquid extraction
Method G	100	Waters Acquity UPLC and Xevo TQ-MS	RSC-Gel 120 C18AqP, 3 µm, 2 × 60 mm	Liquid-liquid extraction using MTBE
Method H	250	Thermo Accela UPLC and Vantage TSQ-MS/MS	Thermo Hypersil Gold C18, 1.9 µm, 2.1x50 mm	Liquid liquid extraction using MTBE
Method J	500	Shimadzu HPLC system and Applied Biosystems Q-Trap 3200 MS	Phenomenex Kinetex Fluorophenyl 2.6 µm, 3.0x100 mm	SPE using J.T Bakker Polarplus Octadecyl (C18) Speedisk

Table 1: **Method characteristics I.** Sample volume, Instrument, Analytical column and sample preparation method per method.

	LLOQ		Calibration range		Internal standard	
	Testost-erone	**Androstene-dione**	**Testost-erone**	**Androstene-dione**	**Testost-erone**	**Androstene-dione**
Method A	0.10	0.10	0.00-26.0	0.00-26.2	$^{13}C_3$-testosterone	$^{13}C_3$-androstenedione

Method B	0.05	0.05	0.03-75.0	0.04-76.0	$^{13}C_3$-testosterone	$^{13}C_3$-androstenedione
Method C	0.10	0.10	0.00-100	0.00-100	$^{2}H_3$-testosterone	$^{2}H_7$-androstenedione
Method D	0.10	0.10	0.10-40.6	0.10-40.6	$^{2}H_3$-testosterone	$^{2}H_3$-testosterone
Method E	0.30	0.30	0.30-50.0	0.30-50.0	$^{2}H_3$-testosterone	$^{2}H_7$-androstenedione
Method F	0.18	0.09	0.35-52.0	0.35-52.0	$^{2}H_5$-testosterone	$^{2}H_7$-androstenedione
Method G	0.17	0.07	0.00-69.3	0.00-34.9.	$^{2}H_8$-17-OH-progesterone	$^{2}H_8$-17-OH-progesterone
Method H	0.20	0.20	0.25-128	0.25-128	$^{2}H_4$-testosterone	$^{2}H_7$-androstenedione
Method J	0.30	0.35	0.00-69.1	0.00-44.9	$^{2}H_8$-17-OH-progesterone	$^{2}H_8$-17-OH-progesterone

Table 2: **Method characteristics II:** Lower Limit of Quantitation, calibration range and internal standard. Concentrations are in nmol/L.

Statistical analysis We calculated the intra-assay CV per analyte of each method using the following formula: CV % = square root{((∑ $(a-b)^2$)/2N)(N/∑$\bar{x}$)}, where ∑ is sum, a and b are the duplicate concentrations of the respective method and analyte, N is total number of duplicates and $\bar{x}$ is the mean analyte concentration of a and b. Recoveries were calculated per method using the following formula: Recovery % = (([analyte] $_{\text{spiked sample}}$ - [analyte]$_{\text{sample without addition}}$)/ [analyte]$_{\text{added}}$)*100%. Recoveries are shown as mean ± SD. Using the mean concentrations calculated from the duplicate measurements of each sample, we performed a Passing-Bablok regression analysis and calculated a Pearson's correlation coefficient to assess the agreement of method B to J with the mean concentration measured by method A. In addition, we calculated the standard deviation (SD) per sample, using the mean concentrations from the duplicate measurements of method B to J. Inter-method variation was calculated using the following formula: CV (%) = SD/[analyte]$_{\text{method A}}$. All statistical analyses were performed using MedCalc 11.6, GraphPad Prism 6 and Microsoft Excel 2010.

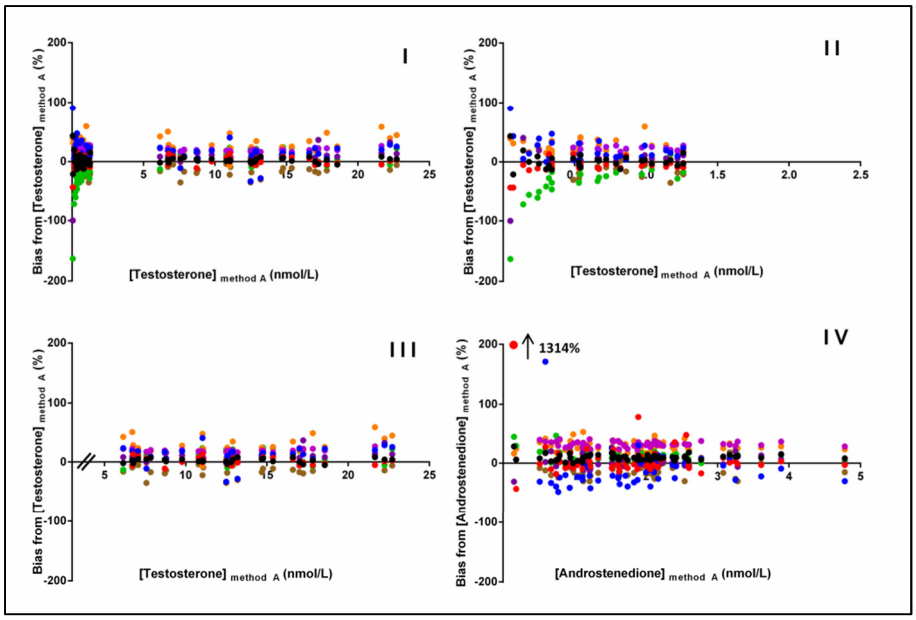


Figure 1: **Bias plots per analyte.** Bias (%) was calculated using the following formula: Bias (%) = (([analyte] method – median [analyte])/ median [analyte]) x 100%. Fig. 1 I-IV show the bias for testosterone (I), for testosterone in females (II), for testosterone in males (III), for androstenedione (IV). Color code: •Method B; • Method C; •Method D; • Method E; • Method F; • Method G; • Method H; • Method J.

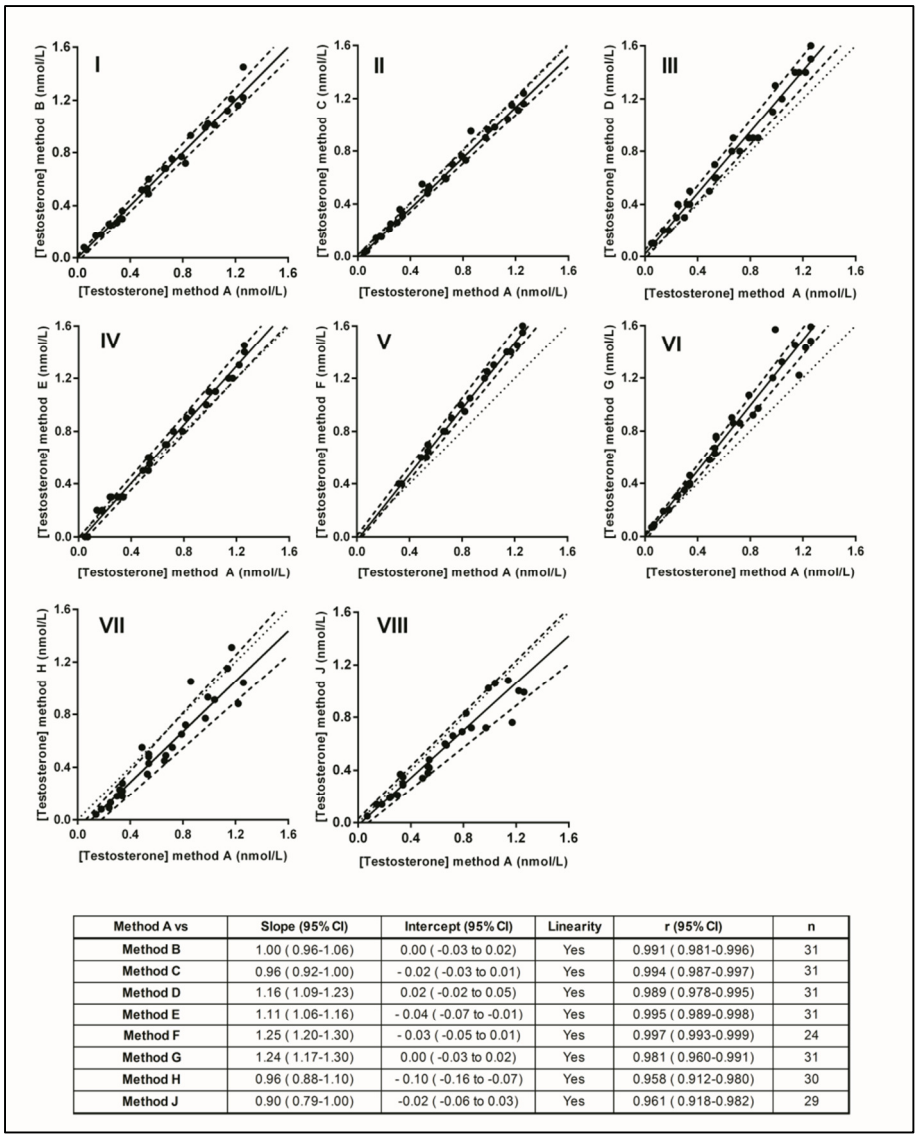
Results

Mean concentrations of testosterone measured by method A were between 0.05-1.26 nmol/L and 6.15-24.44 nmol/L for females (N=31) and males (N=29), respectively. There were no samples with a testosterone concentration between 1.26 and 6.15 nmol/L. Mean androstenedione concentrations measured by method A were between 0.15–4.78 nmol/L. Comparison of the first and second run of method A revealed the following (data are shown as mean (95% C.I.): slope of the regression line was 1.01 (1.00-1.02), intercept was –0.00 (-0.02 to 0.01) and correlation coefficient 0.999 (0.998 to 0.999) for testosterone and the slope of the regression line was 0.99 (0.95-1.03), intercept was –0.02 (-0.07 to 0.04) and correlation coefficient 0.992 (0.986 to 0.995) for androstenedione. For all eight methods studied, the intra-assay CVs, based on the duplicate measurements, were between 3.7 and 16.0% and 0.9 and 5.2%, for testosterone concentrations in females and in males, respectively. Intra-assay CVs for androstenedione were between 1.2 and 9.5%. Intra-assay CVs per method are shown in Table 3.

Method	Testosterone in females	Testosterone in males	Androstenedione
Method A	3.3 %	2.6 %	3.2 %
Method B	7.2 %	4.5 %	3.7 %
Method C	6.5 %	4.1 %	4.3 %
Method D	16.0 %	4.9 %	9.5 %

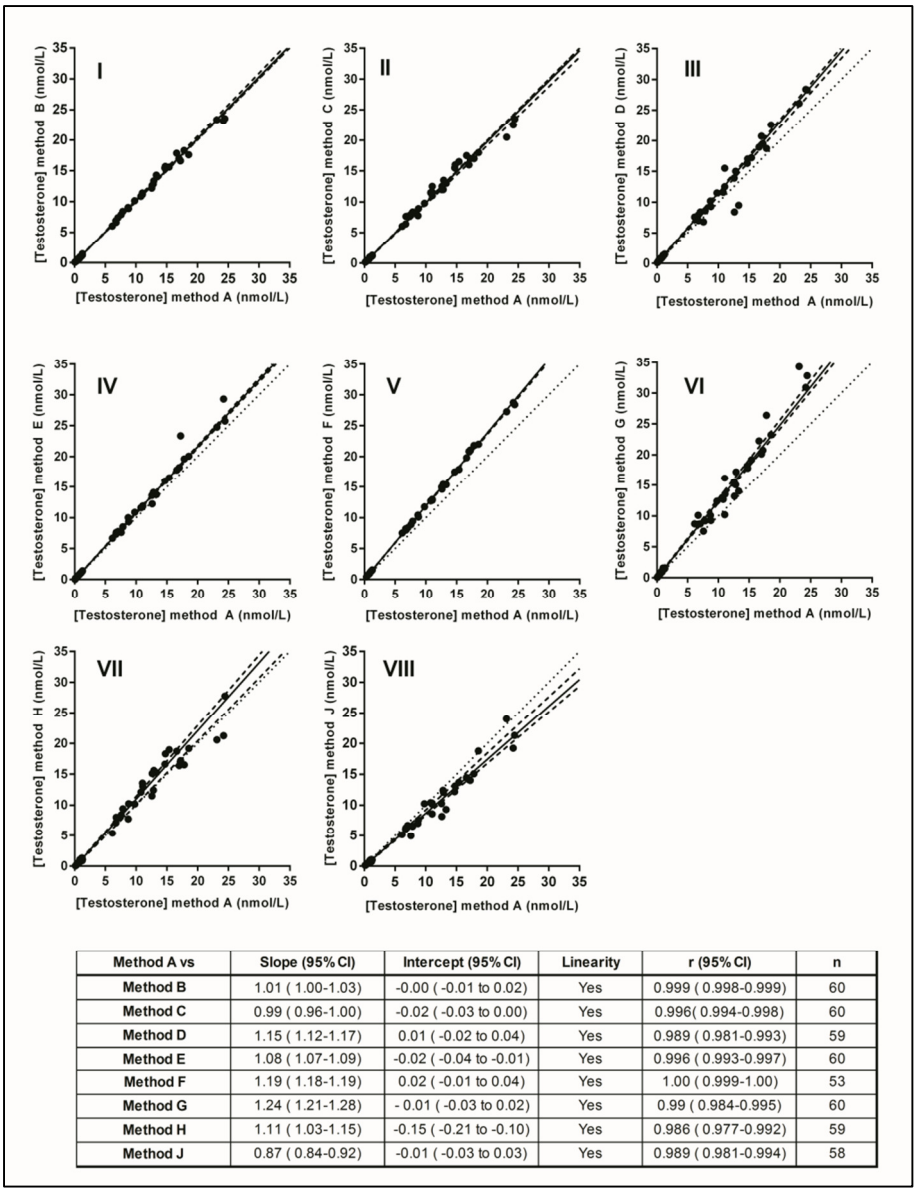
Method E	3.7 %	1.7 %	1.2 %
Method F	4.9 %	0.9 %	1.3 %
Method G	9.3 %	5.2 %	4.3 %
Method H	8.0 %	3.1 %	3.4 %
Method J	6.0 %	3.2 %	5.2 %

Table 3: **Intra-assay CVs per analyte per method.** CVs were calculated from duplicate measurements in one run.



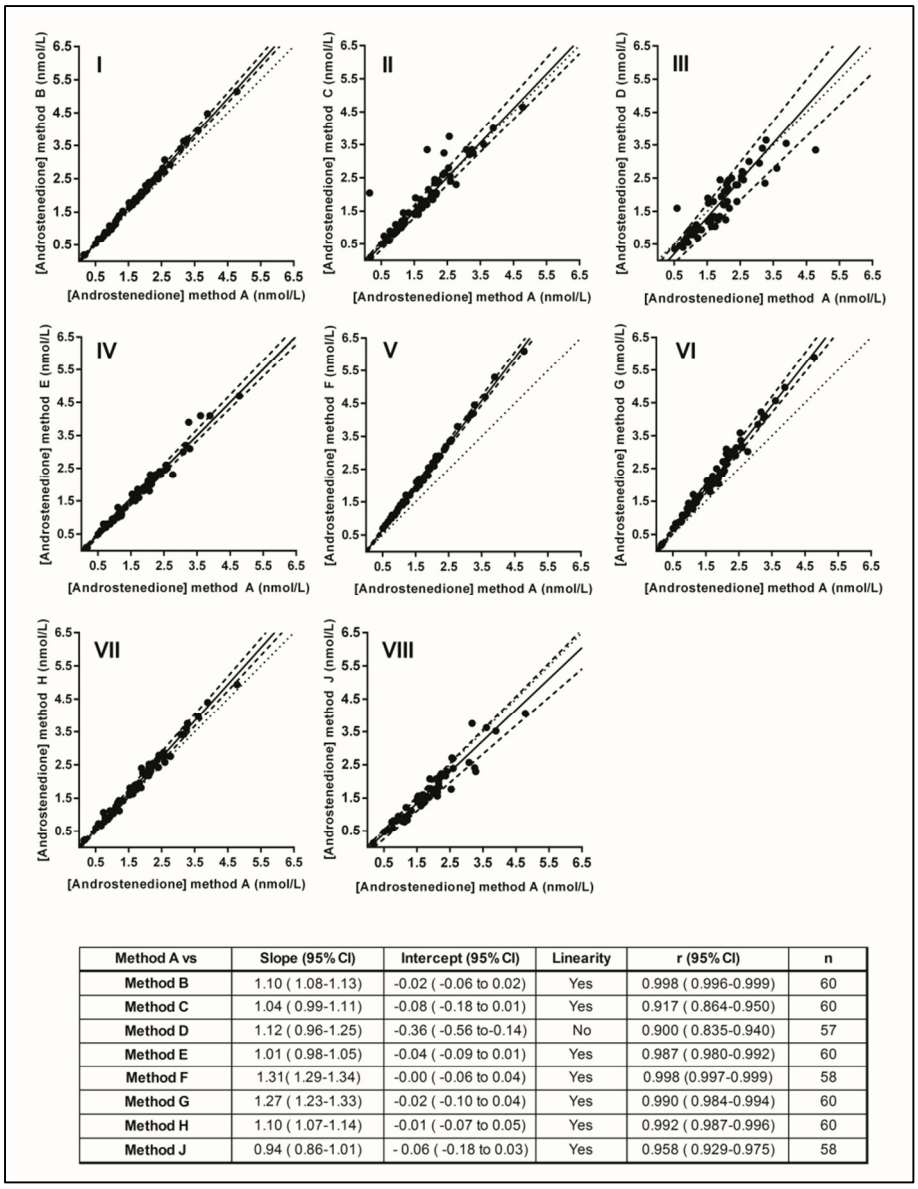
Method A vs	Slope (95% CI)	Intercept (95% CI)	Linearity	r (95% CI)	n
Method B	1.00 (0.96-1.06)	0.00 (-0.03 to 0.02)	Yes	0.991 (0.981-0.996)	31
Method C	0.96 (0.92-1.00)	- 0.02 (-0.03 to 0.01)	Yes	0.994 (0.987-0.997)	31
Method D	1.16 (1.09-1.23)	0.02 (-0.02 to 0.05)	Yes	0.989 (0.978-0.995)	31
Method E	1.11 (1.06-1.16)	- 0.04 (-0.07 to -0.01)	Yes	0.995 (0.989-0.998)	31
Method F	1.25 (1.20-1.30)	- 0.03 (-0.05 to 0.01)	Yes	0.997 (0.993-0.999)	24
Method G	1.24 (1.17-1.30)	0.00 (-0.03 to 0.02)	Yes	0.981 (0.960-0.991)	31
Method H	0.96 (0.88-1.10)	- 0.10 (-0.16 to -0.07)	Yes	0.958 (0.912-0.980)	30
Method J	0.90 (0.79-1.00)	-0.02 (-0.06 to 0.03)	Yes	0.961 (0.918-0.982)	29

Figure 2: **Passing-Bablok regression analyses and Pearson's correlation coefficient for testosterone values in females (below 2 nmol/L).** Testosterone concentration as measured by each of the methods B to J are plotted against the testosterone concentration measured by method A.



Method A vs	Slope (95% CI)	Intercept (95% CI)	Linearity	r (95% CI)	n
Method B	1.01 (1.00-1.03)	-0.00 (-0.01 to 0.02)	Yes	0.999 (0.998-0.999)	60
Method C	0.99 (0.96-1.00)	-0.02 (-0.03 to 0.00)	Yes	0.996(0.994-0.998)	60
Method D	1.15 (1.12-1.17)	0.01 (-0.02 to 0.04)	Yes	0.989 (0.981-0.993)	59
Method E	1.08 (1.07-1.09)	-0.02 (-0.04 to -0.01)	Yes	0.996 (0.993-0.997)	60
Method F	1.19 (1.18-1.19)	0.02 (-0.01 to 0.04)	Yes	1.00 (0.999-1.00)	53
Method G	1.24 (1.21-1.28)	- 0.01 (-0.03 to 0.02)	Yes	0.99 (0.984-0.995)	60
Method H	1.11 (1.03-1.15)	-0.15 (-0.21 to -0.10)	Yes	0.986 (0.977-0.992)	59
Method J	0.87 (0.84-0.92)	-0.01 (-0.03 to 0.03)	Yes	0.989 (0.981-0.994)	58

Figure 3: **Passing-Bablok regression analyses and Pearson's correlation coefficient for all testosterone concentrations.** Testosterone concentration as measured by each of the methods B to J are plotted against the testosterone concentration measured by method A.



Method A vs	Slope (95% CI)	Intercept (95% CI)	Linearity	r (95% CI)	n
Method B	1.10 (1.08-1.13)	-0.02 (-0.06 to 0.02)	Yes	0.998 (0.996-0.999)	60
Method C	1.04 (0.99-1.11)	-0.08 (-0.18 to 0.01)	Yes	0.917 (0.864-0.950)	60
Method D	1.12 (0.96-1.25)	-0.36 (-0.56 to-0.14)	No	0.900 (0.835-0.940)	57
Method E	1.01 (0.98-1.05)	-0.04 (-0.09 to 0.01)	Yes	0.987 (0.980-0.992)	60
Method F	1.31(1.29-1.34)	-0.00 (-0.06 to 0.04)	Yes	0.998 (0.997-0.999)	58
Method G	1.27 (1.23-1.33)	-0.02 (-0.10 to 0.04)	Yes	0.990 (0.984-0.994)	60
Method H	1.10 (1.07-1.14)	-0.01 (-0.07 to 0.05)	Yes	0.992 (0.987-0.996)	60
Method J	0.94 (0.86-1.01)	- 0.06 (-0.18 to 0.03)	Yes	0.958 (0.929-0.975)	58

Figure 4: **Passing-Bablok regression analyses and Pearson's correlation coefficient for androstenedione.** Androstenedione concentration as measured by each of the methods B to J are plotted against the androstenedione concentration measured by method A.

Recoveries for testosterone were between 96 and 111% for all methods, except method D and J showing a mean testosterone recovery of only 58% and 64%, respectively. Recoveries for androstenedione were between 85 and 117% for all methods except method J with a mean androstenedione recovery of 63%. Recoveries per method are shown in Table 4.

	Testosterone	Androstenedione
	n = 6	n = 6
Method A	99± 2 %	87 ± 3 %
Method B	96 ± 5 %	92 ± 5 %
Method C	111 ± 5 %	95 ± 4 %

Method D	58 ± 22 %	89 ± 11 %
Method E	101 ± 7 %	85 ± 6 %
Method F	111 ± 3 %	109 ± 6 %
Method G	102 ± 6 %	117± 10 %
Method H	108 ± 10 %	100± 5 %
Method J	64 ± 7 %	63 ± 4 %

Table 4: **Recoveries per analyte per method.** Recoveries are showed as mean ± SD. Recoveries were calculated using the following formula: Recovery % = (([analyte] $_{\text{spiked sample}}$ - [analyte]$_{\text{sample without addition}}$)/ [analyte]$_{\text{added}}$ *100%.

Bias plots are shown in Figure 1. Passing-Bablok regression analyses as well as Pearson's correlation coefficients are shown in Figures 2-4, for testosterone concentrations in females, all testosterone concentrations and androstenedione, respectively.

The inter-method variation is shown in Figure 5. The mean CV between the eight investigated methods was 19%, 24% and 14% for all testosterone concentrations, testosterone concentrations in females, testosterone concentrations in males, respectively. For androstenedione the inter-method variation was 29%.

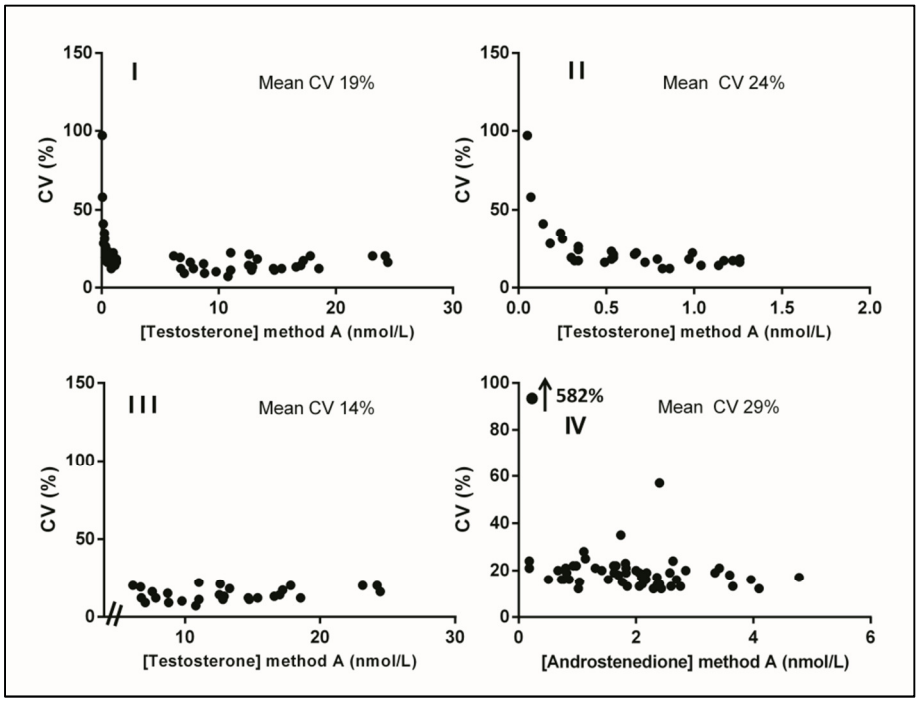


Figure 5: **Inter-method variation based on eight routine LC-MS/MS methods (method B-J).** Inter-method variation was calculated per sample using the formula CV (%) = SD/[analyte] $_{\text{method A}}$. Fig. 5 I-IV show the CVs per sample for testosterone (I), for testosterone values in females (II), for testosterone values in males(III) and for androstenedione (IV).

Discussion

In this study we investigated the agreement and variation between eight routine unpublished LC-MS/MS methods for the simultaneous measurement of testosterone and androstenedione. For testosterone, the agreement of the LC-MS/MS methods investigated is reasonable to good. Four

methods (method B, C, E and H) showed a good agreement (<10% bias) in both the female and male concentration range for testosterone. Four methods (method D, F, G and J) had a structural positive bias when compared to method A of 15%, 22%, 24% and 12%, respectively. This is possibly due to differences in standardization, but could also be caused by matrix effects or cross reactivity. Due to a relatively high variation in the female concentration range, method H and J correlated less well ($r < 0.97$) to method A at lower testosterone levels than other methods. This higher variation was not caused by a higher intra-assay variation as all methods, except method D, had an intra-assay CV below 10% for the entire concentration range. Method D and J reported surprisingly low testosterone recoveries (~60%). These low recoveries cannot be explained by standardization differences, as this difference was not seen in the regression analyses for these methods. Matrix effects might have caused the low recoveries and probably also the relatively poor agreement in the method comparison, as both methods do not use a separate internal standard for testosterone. The testosterone inter-method variation for the currently investigated routine LC-MS/MS methods (24% and 14% in the female and male concentration range) was somewhat poorer when compared to the variation for seven published LC-MS/MS methods (14% and 8% respectively), however better compared to the inter-method variation for seven currently used immunoassays (73% and 12% respectively) [13].

For androstenedione the agreement between four of the investigated methods (method B, C, E and J) and method A was good (bias <10%). Four methods (method D, F, G and H) seem to have a structural positive bias around 12%, 31%, 27% and 11% when compared to method A, possibly caused by a difference in the standardization. Due to a relatively high variation (method D and J) and outliers (method C), these methods correlated less well ($r < 0.96$) to method A than other methods. All methods had an acceptable intra-assay CV for androstenedione, below 10% for the entire concentration range. The lower recovery (64%) of method J cannot be explained by standardization differences, as this difference was not seen in the regression analysis, but might originate from matrix interferences, due to the fact that no separate internal standard for androstenedione was used in method J. Method C showed four outliers for androstenedione. These outliers were caused by an interference in the androstenedione quantifier trace. Further research into the origin of this interference as well as its elimination is currently undertaken. The inter-method variation for androstenedione (29%) was higher than the inter-method variation observed for testosterone. Four samples outlied in method C and caused a higher mean inter-method variation. When these four samples were excluded the mean inter-method variation dropped to 19%, a variation comparable to the one found for testosterone methods. This is still somewhat higher than the variation observed between the seven published LC-MS/MS assays compared earlier (12%) [13]. Unfortunately, no published data are available on the variation between immunoassays for androstenedione.

Although a testosterone reference method was not included in this study, method A has been indirectly compared to a reference method and showed a good agreement with this reference method [12] as well as with six published LC-MS/MS methods for testosterone [13]. Notably, we

found a high agreement between several methods in the present study, suggesting that these methods measure close to the true testosterone concentrations. Restandardisation and/or the elimination of the high variation might bring the other investigated methods to this point as well. For androstenedione a reference method is not available; therefore trueness of the methods in this study cannot be estimated. Despite this limitation, the conclusion that we do observe a difference in standardisation between the LC-MS/MS for androstenedione appears to be justified.

As we observed a better agreement in the earlier study concerning published LC-MS/MS method compared to what we found in the current study, we speculate that a peer-review process stimulates the efforts put in method development and validation. Furthermore, the current study stimulated some of the participants to reveal the origin and solve the shortcomings found in this method commparison. Some of the problems found in this study might have their origin in (unknown) changes in a previously well validated method. An earlier study showed that changes, like change in vials used, in the method may cause problems in a previously well performing method [13]. This underlines the need not only for initial validation but also for the continuous monitoring of LC-MS/MS methods, for example by participation in internal and external quality assessment schemes as well as more elaborate method comparison studies such as the present investigation.

In conclusion, most of the investigated routine LC-MS/MS methods for testosterone and androstenedione showed a reasonable to good agreement. This study demonstrates that standardization and variation need improvement in some of the methods. The better inter-method variation between LC-MS/MS assays compared to immunoassays for testosterone, supports the JCEM-policy to accept manuscripts reporting testosterone in the lower concentration range only when these are measured by mass spectrometry based methods [8]. However, the differences in standardisation and variability found between the various LC-MS/MS methods, suggest that such a policy alone is not an assurance for high quality sex steroid measurement results. A future statement concerning hormone measurements should probably not include the requirement for LC-MS/MS assays as such, but emphazise the requirement for accurate assays of any analytical method, and above all the demand from authors to provide data that prove that the method used underwent a thorough validation and is under continuous quality control.

References

1. Taieb J, Mathian B, Millot F et al. Testosterone measured by 10 immunoassays and by isotope-dilution gas chromatography-mass spectrometry in sera from 116 men, women, and children. Clin Chem 2003; 49:1381-95.

2. Herold DA, Fitzgerald RL. Immunoassays for testosterone in women: better than a guess? Clin Chem 2003; 49:1250-1.

3. Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H. Position statement: Utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. J Clin Endocrinol Metab 2007; 92:405-13.

4. Rosner W, Vesper H. Toward excellence in testosterone testing: a consensus statement. J Clin Endocrinol Metab 2010; 95:4542-8.

5. Groenestege WMT, Bui HN, ten Kate J et al. Accuracy of first and second generation testosterone assays and improvement through sample extraction. Clin Chem 2012; 58:1154-6.

6. Bui HN, Sluss PM, Blincko S, Knol DL, Blankenstein MA, Heijboer AC. Dynamics of serum testosterone during the menstrual cycle evaluated by daily measurements with an ID-LC-MS/MS method and a 2nd generation automated immunoassay. Steroids 2013; 78:96-101.

7. Ackermans MT, Endert E. LC-MS/MS in endocrinology: what is the profit of the last 5 years? Bioanalysis 2014; 6:43-57.

8. Handelsman DJ, Wartofsky L. Requirement for mass spectrometry sex steroid assays in the Journal of Clinical Endocrinology and Metabolism. J Clin Endocrinol Metab 2013; 98:3971-3.

9. Thienpont LM, Van Uytfanghe K, Blincko S et al. State-of-the-art of serum testosterone measurement by isotope dilution-liquid chromatography-tandem mass spectrometry. Clin Chem 2008; 54:1290-7.

10. Vesper HW, Bhasin S, Wang C et al. Interlaboratory comparison study of serum total testosterone measurements performed by mass spectrometry methods. Steroids 2009; 74:498-503.

11. Legro RS, Schlaff WD, Diamond MP et al. Total testosterone assays in women with polycystic ovary syndrome: precision and correlation with hirsutism. J Clin Endocrinol Metab 2010; 95:5305-13.

12. Büttler RM, Martens F, Kushnir MM, Ackermans MT, Blankenstein MA, Heijboer AC. Simultaneous measurement of testosterone, androstenedione and dehydroepiandrosterone (DHEA) in serum and plasma using Isotope-Dilution 2-Dimension Ultra High Performance Liquid-Chromatography Tandem Mass Spectrometry (ID-LC-MS/MS). Clin Chim Acta 2015; 438:157-9.

13. Büttler RM, Martens F, Fanelli F et al. Comparison of 7 Published LC-MS/MS Methods for the Simultaneous Measurement of Testosterone, Androstenedione, and Dehydroepiandrosterone in Serum. Clin Chem 2015; 61:1475-83.